

Association for Information Systems AIS Electronic Library (AISeL)

PACIS 2008 Proceedings

Pacific Asia Conference on Information Systems
(PACIS)

July 2008

Exploiting the Rich Document Structures and Network Topology of Legal Information Systems

John Ellis Murphy

University of Technology Sydney, jemurphy@it.uts.edu.au

Robert Steele

University of Technology Sydney, rsteale@it.uts.edu.au

Rita Shen

University of Technology Sydney, ritashen@it.uts.edu.au

Follow this and additional works at: <http://aisel.aisnet.org/pacis2008>

Recommended Citation

Murphy, John Ellis; Steele, Robert; and Shen, Rita, "Exploiting the Rich Document Structures and Network Topology of Legal Information Systems" (2008). *PACIS 2008 Proceedings*. 234.
<http://aisel.aisnet.org/pacis2008/234>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2008 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

EXPLOITING THE RICH DOCUMENT STRUCTURES AND NETWORK TOPOLOGY OF LEGAL INFORMATION SYSTEMS

John Murphy, Department of Computer Science, University of Technology Sydney, PO Box 123 Sydney, 2007, Australia, jemurphy@it.uts.edu.au

Robert Steele, Department of Computer Science, University of Technology Sydney, PO Box 123 Sydney, 2007, Australia, rsteele@it.uts.edu.au

Rita Shen, Department of Computer Science, University of Technology Sydney, PO Box 123 Sydney, 2007, Australia, ritashen@it.uts.edu.au

Abstract

The move toward on-line publishing of legal documents in the public domain has made extra information available by way of mark-up for electronic publishing. Legal Concepts that were once implicit in documents are now explicit and further inferences can now be drawn through computation however the quality and extent of this information varies. This paper discusses the application of web techniques to the legal system for the ranking of related legal concepts. We demonstrate how these techniques can be adapted and applied to the legal system and how strategies can be used to optimize the best combination of techniques depending on the quality of the source information. Courts and legislatures create and distribute laws through documents that serve as the primary sources of law within legal systems. With the increase in on-line publishing, these documents are being marked-up with greater quantities of information particularly with links to other related sources within the legal system.

Keywords: Legal Information Systems, Network Topologies, Conceptual Information Retrieval.

1 INTRODUCTION

An increasingly important aspect in the growth of industries in a global business environment is the laws that govern business activities in different jurisdictions. The ability to conduct business within regulatory frameworks requires an understanding of the complex interrelated network of legislation and case law within different legal systems. This is an area in which Information Systems can provide solutions to assist the understanding of this interconnected network of laws and regulation.

Information systems have evolved to deal with problems that resemble or can be framed in terms of networks with much research into social networks and link based Page ranking algorithms for the World Wide Web. The regulatory environments of legislation and case law are composed of documents that are well structured and contain references to other legal sources in the form of case citations, legislative references and digest references. The legal system can thus be viewed as a web of interrelated legal rules and legal concepts contained in documents and linked by citation references.

This paper describes an approach to assigning authoritative values to related legal issues. This approach requires no additional information and minimal change to the existing systems and may have applications including enhanced search engines, research oriented visualization tools for understanding legal relationships, tools to assist legal drafting and expert legal systems.

2 BACKGROUND

Primary sources of law from courts and parliaments are embodied in documents in the form of case reports and legislation. In order to understand the law, a practitioner must be aware of the relationships between different cases, cases and legislation, and the relationships between different pieces of legislation (Hafner, 1987). The legal principles underpinning the law, are often distributed across multiple legislative and Case report documents. In addition to being distributed across documents, legal principles evolve overtime and therefore the temporal aspects of the documents must be considered such as what provisions have been repealed or what cases have been overturned and what new decisions have been made in relation to the legal principles.

Case Law incorporates the set of laws evolved through the court systems that is grounded in the principles of precedent and case authority. A legal case is judged according to the legal principles laid down in previous cases. Legal advocates often cite other cases as authority for the arguments they present and judges cite cases relied on when making their decision. This interconnection of cases through citations forms a network of Case Law as described by (Zhang & Koppaka, 2006) that practitioners must understand and use in order to carry out their duties within a common law legal system. The understanding of the implications of the related legal principles within Case Law requires expert judgement and the capacity to understand the structure and relationships within the system.

3 NETWORK TOPOLOGY OF A LEGAL INFORMATION SYSTEM

Research into the network topology of the web has produced models and algorithms for determining the importance and relevance of web pages for the purpose of retrieving information. The link based Page Ranking algorithms of (Page et al., 1998) and (Kleinberg, 1999) have yielded a greater understanding of how the network of documents on the web is formed over time and how patterns reflecting common topics and authority can be elicited from the network. The World Wide Web can be interpreted as a networked system of documents connected by hyper-links. The principles of link based ranking can be adapted and applied to the system of linked cases and legislation to exploit the network topology found in the internal structure and citations of legal documents.

The network topology of a legal information system is formed by the documented relationships between the institutions, roles and processes that compose the legal system. These components include

the courts, parliaments, judges, lawyers, hearings, appeals, judgements, case reports and statutes. The legal system has evolved a number of aspects that give formality and structure to the legal information system of case law and legislative documents. It is these aspects of the legal system that distinguish it from other less formal domains that enable it to be exploited for purposes such as information retrieval, document management and the exploration and understanding of legal issues.

3.1 Exploitable Aspects of the Legal System

There are a number of important differences between a legal system and the World Wide Web that can be exploited to enhance the link based ranking approaches used for the web. These differences relate to the nature of the legal community such as the language, the formal hierarchy and commonality of purpose. Other important distinctions relate to the production of the documents such as the content, structure and classification mechanisms and there is a final distinction that revolves around the system of institutions within the legal system that follow a strict procedural pattern. These exploitable distinctions include:

1. **Authority Context.** The legal community can be regarded as an homogeneous group for the purpose of legal concept recognition and authority. Although differences in the degree of authority occur between individuals roles and institutions in the legal system, legal practitioners are interested in the same legal concepts. On the other hand, the web can have diverse authorities of opinion simply because participants belonging to a different community are unlikely to be interested in a topic that has strong endorsement by participants in another community of interest.
2. **Link Semantics.** Citation and reference links in legal cases and legislation have much richer meaning than the vague notion of user endorsement found on the web. Web links often contain non-endorsing links such as advertising or dynamic links that negatively impact upon the concept of endorsement and thus reduce the reliability of the links. Legal cases and legislation contain well defined links with richer semantics such as court rulings that have been followed or overturned and judgements agreed with or disagreed with. These links not only have greater certainty but provide more meaningful information.
3. **Document Structure.** Legislative and legal case documents are well structured documents that infer other relationships beyond the semantic links between web documents inferred from page co-occurrence and link text. Case references occurring in certain parts of a case document have different meaning because of the context of their location in the document. For example a citation in the held section of a case document holds greater significance than a citation found in procedural matters or general argument.
4. **Participant Roles.** The roles of participants in the court system can be used to convey extra meaning to a citation. A citation given by a legal advocate in argument is not as strong as a citation given by a judge in the reasons for a decision.
5. **Multiple Topics.** A legal Case is often associated with more than one legal issue. Although the different issues are closely related to each other via the facts of the case and the court proceedings, they may not be closely related in terms of recognized common legal principles within the law. It is therefore important to separate the legal issues dealt with in a court case when calculating the authority of a case. Pages on the web by contrast are usually interpreted as addressing a single topic for the purposes of ranking the authority of pages.
6. **Term Ambiguity.** The legal system generally moves away from ambiguity and toward a common understanding through the language of law which is more precise than non legal language and through the system of recognised authorities in the form of appellate

courts that determine the accepted meaning of terminology and legal concepts. By contrast, the web uses common language with its inherent ambiguities and has no recognized central arbiter of the meaning of terms.

These aspects of the legal system provide greater certainty and structure upon which information systems can be built. The law has evolved a paper based information system that includes documents such as case reports, legal digests, thesauri and statutes. The recent moves toward conversion of these documents to electronic form has yielded sources of rich structure and relationships that can be exploited by legal information systems.

4 RICH DESCRIPTION MODEL OF THE CASE REPORT

Case reports and statutes are important documents that embody the laws produced by the courts and parliaments of the legal system. These documents through their content, structure and references convey the complex information that constitutes the law. Central to the exploitation of the enriched links between legal concepts is the development of a model of the case report document. The case report is a well structured document used within the legal system to report the legal information produced in court proceedings and conveys the common law legal principles and interpretations of statutes required to fully understand the law. Our model categorizes the content of a case report into five types of information.

- Institutional and procedural information concerning the progression of a matter through the Court system. This information includes the roles of the participants in the matter such as the presiding judge, and the advocates of the parties involved in the dispute. It further includes information such as the type of court, the orders made and whether the case is being heard on appeal.
- The main body of the the case report which the judges make their determinations about the case and explain their decisions. This is portion of the case report document that contains the details of the arguments of the case in specific legal paragraphs.
- Information that identifies and summarizes the important principles considered within the body of the judgement. This part of the document called the holdings is produced by expert case reporters as part of an official court publication. This is the portion of the case report in which legal experts publish their review of the important legal issues considered in the judgement and relate those issues to external references such as legal digests. The holding portion of a case report is well structured with internal references that link the parts of the judgement that deal with the important legal issues considered in the case.
- The parts of the document that categorize the case by different criteria. These parts include
 - categorization by common legal words and phrases in the judgement that are recognized as key terms in the legal system. These words and phrases are identified by legal experts and can be interpreted as tokens that mark well recognized legal issues within the law.
 - a set of key words known as legal catch words that relate a case to a commonly accepted external hierarchy of legal terms. These terms are derived from a common published legal digest that is used to define and categorize the law. These keywords are also used in the case report to identify the important legal issues.
 - classifications that contain extra information about references such as indexes of cases and legislation considered and commented upon in the judgement.
- References to other legal sources such as cases, legislation and digests. These explicit references are used as the foundation of the legal network topology.

Case reports also contain implicit higher level legal concepts that can be derived from the explicitly marked-up concepts in the case report. An example of developing higher level legal concepts from specific concepts explicit in the document is the Considered Case concept that is implied by the citation link identified in the case report extract in figure 1 as “AAA v BBB (1998) 23 FAM LR 716, affirmed”. This is the citation of case that the judge took into consideration when deciding the KR v KR case, specifically Affirming the previous decision. The Considered Case is a higher level legal

concept and a judicial mechanisms that supports the concept of Case Authority in common law. Case Authority can be interpreted as the application of cumulative weightings of importance to a case over time using the decisions of subsequent judgements that have considered the case and Followed, Not Followed, Reversed or Applied the case rulings.

<p>KR v KR [2005] NSWCA 411 Court of Appeal: Hodgson JA, Bryson JA and Hunt A-JA 22 September, 25 November 2005</p> <p><i>Family Law --- Defacto relationships --- Adjustment of property interests --- Relevant Property --- Property (Relationships Act 1984, s 20.</i></p> <p><i>Held:</i> (1) The exercise of jurisdiction under s 20 ... (2)The court may take a global view ... AAA v BBB (1998) 23 Fam LR 716, affirmed. Green v Robinson (1995) 36 NSWLR 96, not followed.</p>
--

Figure 1. Example Case Report Document.

The extract in figure 1 is taken from the Head Note portion of the case report and shows a structured mix of judicial case management information and information concerning legal principles. In addition to the Judgement, the case report incorporates elements of judicial process such as hearing dates, parties and their advocates, the presiding judge, the type of court, descriptions of the nature of the case and cases cited. Furthermore the Case Report includes more complex parts such as the “Catch Words” block that is a hierarchical categorization of key terms derived from Legal Digests and a part that summarizes the important legal issues dealt with by the case. Mixed with the structured parts of the document are links to external unifying sources such as common legal phrases and terms from Legal Thesauri and links to relevant legislation.

5 LINK BASED CASE RANKING APPROACH

The proposed Link Based Case Ranking approach calculates rankings for the authority of legal cases in respect to the legal issues covered by the case. The approach exploits the rich document structure, meta-data and references found explicitly and implicitly in legal documentation. The approach involves a number of stages that derive new relationships between the legal issues expressed within documents and adds these relationships as new links to the basic Network Topology formed by the case citations and references.

Page ranking algorithms for the web rely on a simple network topology of pages as nodes and hyperlinks as links. The page is generally interpreted as representing content about a single topic and the links between pages are interpreted as user endorsement of the page (Nie et al., 2005). The nature of the legal system and the rich structure of its documents allows for the extension of the interpretations of the nodes and links in the legal network topology. Page Ranking algorithms also assume that links have the same value to begin and recursively adjust the weightings using algorithms based on a count of links into and out of a node.

It is usual for courts to consider a number of legal issues that are regarded as unconnected or loosely connected in law when dealing with a matter. These legal issues are the concepts that legal practitioners must relate to the facts of a legal matter to understand how the law applies to a case. It is

the legal issue and not the case that is the unit of understanding. For this reason, the type of node that is most significant in the legal network topology is the legal issue. The case report explicitly identifies the important legal issues addressed by the court in well defined parts of the document as well as providing summaries and details of the issues in identifiable legal paragraphs. Each of these structured parts of the case report is interpreted as a type of node that can be connected to convey extra information relating to a specific legal issue.

The rich structure of Case Reports provide extra information that can be exploited by incorporating the information into the Network Topology as different types of nodes and links. The type of information ranges from simple document structures such as the legal paragraph that represents a legal issue through to citations of judicially considered cases that apply the previous decision of a superior court. The introduction of new nodes allows for new types of links in the topology that provide more information.

The objective of the link based case ranking approach is to calculate weightings for the legal issues contained in a case so that an authority ranking can be assigned for each specific legal issue addressed by the case. The calculation is based on the structure of the legal network topology and relies on the presence of connections between the legal issue nodes to propagate the weightings through the network. Legal issues are explicitly defined but are not always explicitly connected in case report documents. The introduction of new types of nodes and links builds paths between the nodes that can later be used to propagate weightings to the nodes through these extra connections (Haveliwala, 2002). The process of building connections starts with the basic topology formed by the explicit citation links between cases. The topology is expanded by the addition of new nodes and connections based on the structures and meta-data explicitly marked-up in the document text.

There are a number of important difference between this approach to link based ranking and those of the web. The first difference is that web approaches calculate weightings for nodes using a single type of link and a single type of node. The second difference is that links in the legal topology have predefined weightings assigned to them that affect the node values propagated through those links. Lastly the nodes in the web approaches are assigned equal values before the propagation of weightings across the network. It is the nature of the legal system and the structure of its documents that enable the inclusion of this extra information within the network topology.

5.1 Process of Identifying or Creating Nodes and Links

The process of identifying legal issue nodes involves a number of steps beginning with the identification of the basic structures of the case report document and its explicit references. The process then progresses to the identification of legal concepts at progressively higher levels of abstraction using information found in the mark-up of case report documents. At each iteration of the process, further inference are made to either identify or create the nodes and links in the network topology.

The iterative approach of starting with simple concepts and moving toward the more complex is taken because the quality of the source data is not always reliable and this approach maximizes the amount of information that can be formalized into the model. The five steps of the process are:

- Identify the explicit links between cases and statutes in the form of case citations and legislative references. This part of the process often requires minimal parsing of textual references. The need for parsing is diminishing as the use of formally defined Medium Neutral Citations (MNC) is becoming increasingly common. Figure 1 includes an example MNC.
- Identify and extract any meta-data that is a property of the explicit links. An example of this type of meta-data is found with case citations within the held section of a case report. Figure1 includes two example case citations in the held section that include additional information about how the case was used in the judgement. The first case was Affirmed and the second Not

Followed. This meta-data can also be implied in circumstances where the link is included in a collection or situated in a particular part of a document. An example is a reference to a statute that is included in a list of statutes judicially considered for the case.

- Identify and extract implicit knowledge associated with the explicit links. This information is usually inferred from content surrounding the link although some case report documents explicitly mark-up this information. An example of this type of information is the role of the party citing a case reference. A case cited by a legal advocate arguing for a favourable interpretation on some legal issue will hold lesser authority to a case cited by a judge in the reasons for the decision. Some judgement needs to be exercised to determine if this information should be included in the network topology as another node and link or as an attribute of the link that can later be used in the calculation of a nodes weighted value. If the information can be used to establish a relationship between legal issues then it is worth including as a link and node in the topology.
- Identify information implicit in the document structure that can be used to infer links between existing nodes in the network topology. The simplest example of this type of information is the occurrence of two or more citations in a legal paragraph. Co-location is a well known method of establishing relationships between concepts. This principle can be used to inject an implied connection into the topology.
- Infer information from the document structure and meta-data at the legal concept level. It is similar to using the document structure to infer information however it relies on known higher level legal knowledge that can be inferred by combining multiple pieces of information from the document. This type of information is difficult to identify but can be inferred. A good example is the legal concept of Affirmed on Appeal. This is a court case that has been Appealed and Affirmed in the same hearing.

The legal network topology begins to emerge as the links and nodes are identified or created. The complexity of the topology is dependent on the quality of the source legal documents used in the process and the more information available the more complex the model. As this model grows it is important to make intelligent choices to prevent the model from becoming too complex. The choices to include or exclude nodes and links from the topology are determined by the assigned values of the nodes and links informed by experts. These values are later used in the propagation of weightings through the network.

5.2 Process of Assigning Weighted Values

The shape of the topology we wish to create is composed of nodes representing legal issues and connections that represent similarity or equality of legal concept. These representations rarely exist in the explicit network topology formed from the case report and legislative documents however it is possible to achieve a network topology that includes an approximation of this structure by introducing other types of nodes and link. Most explicit links are between different cases, cases and legislation and legal issues and cases where the citation exists in a legal paragraph or a part of the case report that defines a single legal issue. This approach builds up a network of different types of nodes and links however the goal is to identify and link legal issue as demonstrated in the example legal network topology in figure 2.

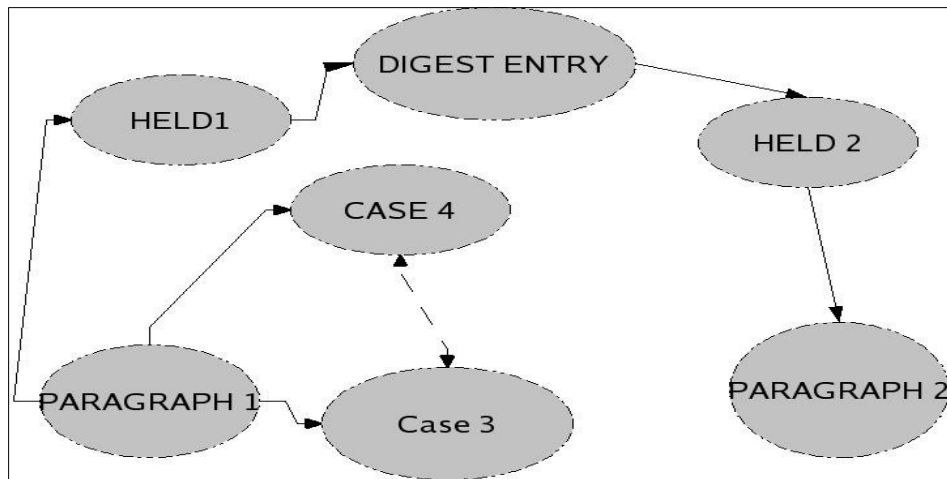


Figure 2. *An Example Fragment of a Legal Network Topology.*

The example in Figure 2 demonstrates the use of external digest entries and the internal structure of the case report to create further connections. The legal issue contained in PARAGRAPH1 is identified and referenced by an entry in the case reports held section HELD1. This held entry contains either a digest reference or digest keywords entry that is shared by the held section of another case report HELD2 in turn references the legal issue dealt with in legal paragraph PARAGRAPH2. The link between case reports CASE3 and CASE4 is a derived link via the co-location of their case citations within PARAGRAPH1. This is one method by which an indirect connection can be made between legal issues in different case report documents.

The information gathered from the five stages of the concept identification process is used to create nodes and links in the network topology model. This process involves the allocation of values to links and nodes based on the type of concept they represent. The process of assigning weighted values is determined by the information available in the source document and informed by prior expert knowledge of the structures, meta-data and relations between them.

A number of different combinations of relationships can be achieved between legal concepts, document structures and meta-data. Each combination relies on the production of weightings or biases using expert knowledge of the concepts in the legal system. Some example of these combinations are:

1. The citation of another case within the body of a judgement by an advocate of a party to the dispute. The extra role based information would enable the assigning of a lower ranking than a citation given by a judge in the decision.
2. A pair of citations co-located in the same part of the held section that summarizes important legal points will yield important information about the common important legal issue that is the basis of the relationship between the cited cases.
3. A case cited by a judge and judicially considered contains additional information as the authority of the previous case. The type of information recorded in a case report indicates if another case has been followed, not followed or applied. This additional information significantly changes the character of the citation link.

The final part of the process is the propagation of values along the network to yield weightings for legal issues. Legal issues are not specifically defined in case reports however reasonable boundaries are provided within the document structure of a case report to mark the legal issues. Demarcation of Legal issues occurs in the held section, catchword block and paragraph. The held section and

catchword blocks are lists in which each item pertains to one legal issue. Legal Paragraphs provide another boundary in the form of a many to one relation between legal paragraphs legal issues. A legal paragraph will not however represent more than one legal issue Figure 1.

The propagation of weightings is carried out in a manner similar to the PageRank algorithm described by (Page et al., 1998). Page's approach uses an algorithm that assigns a weighting to web pages based on both the number of pages that link to it and the weighted authority of those pages. In this way, authoritative pages propagate their authority to pages to which they link. Page's algorithms assist in the identification of pages related to the same topic and to strongly endorsed or authoritative pages. The approach does not however distinguish user community context, consider that pages link to different topics and consider that a page may endorse some links but not endorse others.

The most significant difference between Page's algorithm and our approach is that the nodes are not assumed to hold an equal weighting and the distribution to the next set of nodes is not equally divided among the target nodes but aggregated according to the value of the connections.

5.3 Potential Applications of the Link Based Case Ranking Approach

The link based case ranking approach has potential applications for research aids that enable the navigation of related legal issues once a specific issue has been identified. It also has applications for legal information retrieval systems, enhanced conceptual search engines, expert legal systems, drafting tools, teaching aids for law students and system for the navigation and understanding of legal issues for legal researchers and law students.

6 CONCLUSION

The link based approach to case ranking exploits the formal nature of the legal system and the structure and meta-data of its documents to link related legal issues across disparate documents. This paper describes how this approach uses legal documents to build network topologies and the processes for the identification and creation of links and nodes from the explicit meta-data, document structure and implied legal concepts within case reports. We also describe the process of assigning values to the network of legal concepts for the propagation of weightings to give an authoritative weighting to connected legal issues. Legal information systems are particularly well suited to this approach however it can have application to any domain with formal processes and well structured documents.

References

- Hafner, C. D. (1987) Conceptual organization of case law knowledge bases ICAIL '87: Proceedings of the 1st international conference on Artificial intelligence and law, ACM Press, 35-42
- Haveliwala, T. H. (2002) Topic-sensitive PageRank WWW '02: Proceedings of the 11th international conference on World Wide Web, ACM, 517-526
- Kleinberg, J. (1999) Authorative Sources in a Hyperlinked Environment Journal of the ACM, 46, 604-662
- Nie, Z.; Zhang, Y.; Wen, J. & Ma, W. (2005) Object-level ranking: bringing order to Web objects WWW '05: Proceedings of the 14th international conference on World Wide Web, ACM, 567-574
- Page, L.; Brin, S.; Motwani, R. & T. Winograd (1998) The PageRank Citation Ranking: Bringing order to the web Stanford University
- Zhang, P. & Koppaka, L. (2007) Semantics-based legal citation network ICAIL, 123-130